ANN for neutrino event Selection

- Performance of an ANN, (ANN3) constructed for neutrino event selection, on the 82 event sample selected from period 1 as neutrino interactions and on the new strip files of period 1.

- Explanation of the deviations from what expected.

- Construction and performance (on the 82 event sample and the new strip files) of a second different ANN (ANN4) as an attempt to account for possible changes on the selection criteria for both “signal” and “background” events.

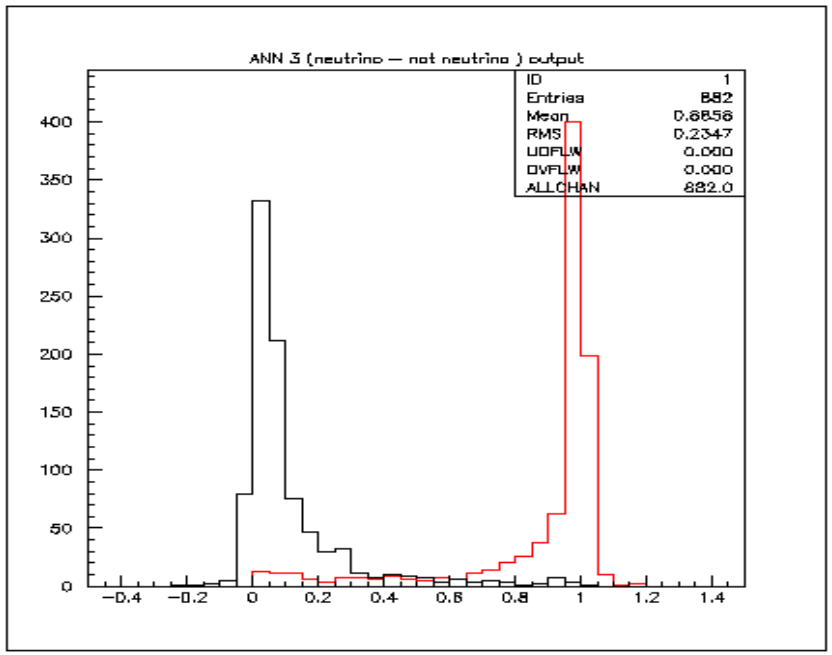
- Results and conclusions.

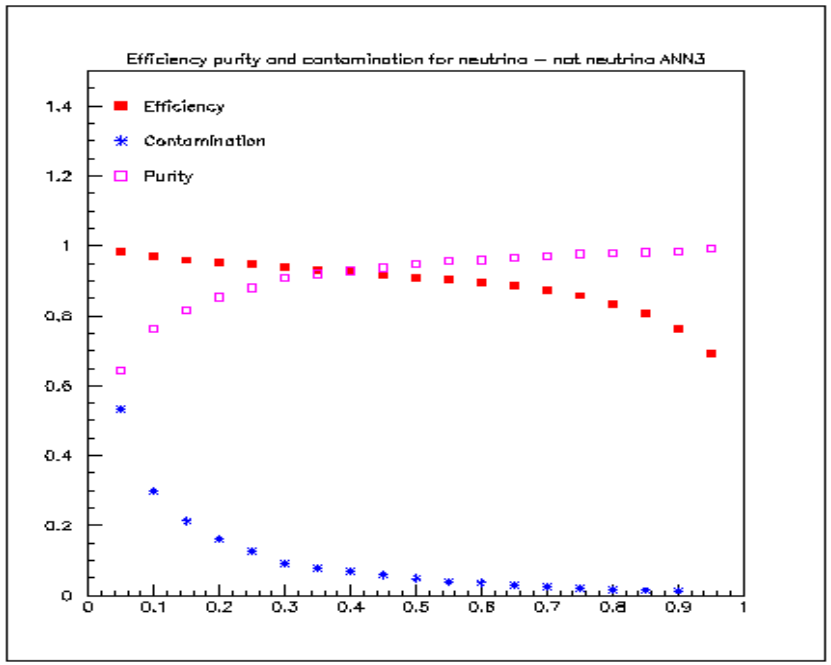
First ANN (ANN3)
Training set - Input Variables

- **Training Set :** We are limited to use as **"signal" training set the 887 neutrino events** and as **"background" training set events randomly selected from OLD nustrip files.**

- **Input variables :**
  - TDC value differences T3-T2,T2-T1,T3-T1
  - Calorimeter energy along y=0 and $|x| > 100$ cm
  - Number of SF, DC, VDC, MID hits,
  - Total Pulse height, % of SF hits in Stations 1 2 3 & 4
  - Total Energy in the EMCAL, number of Clusters, Average Cluster Energy,mean angle of clusters with respect to the z-axis from the interaction vertex
  - Number of SF lines, DC tracks, Final tracks
  - Number of vertices, vertex in or out the emulsion module

First ANN (ANN3)
Output probability distributions

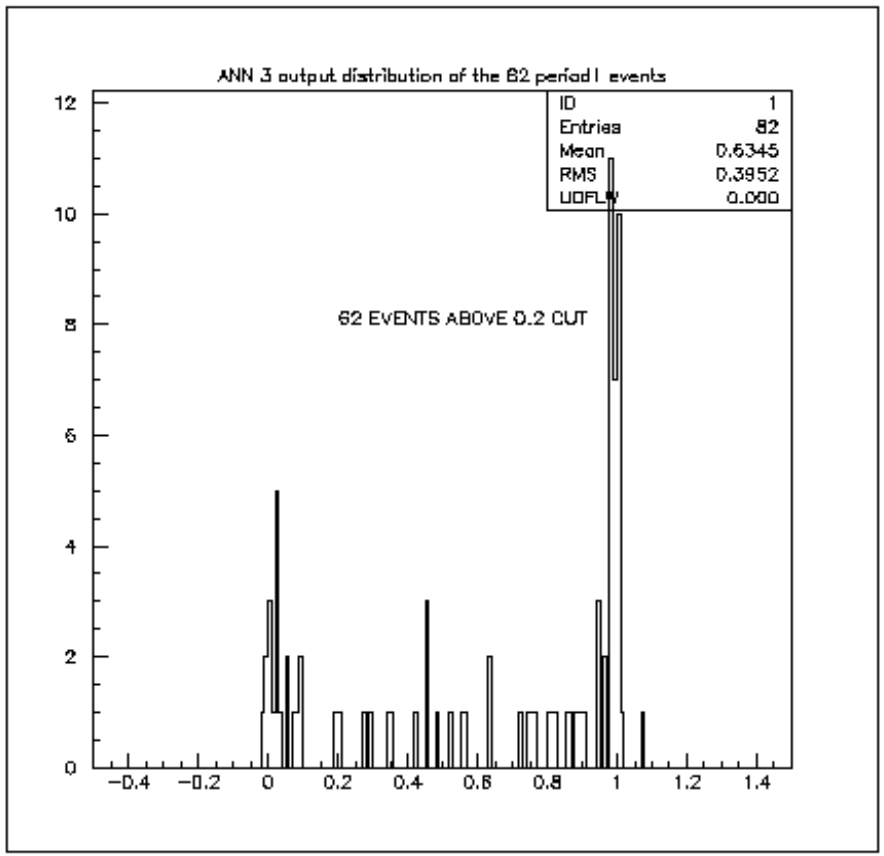




- With a **cut @ 0.2** in the network output function we select :

SIGNAL with **efficiency : 95 % purity 85% contamination 16%**

First ANN (ANN3)
Output distribution on 82 period 1 events

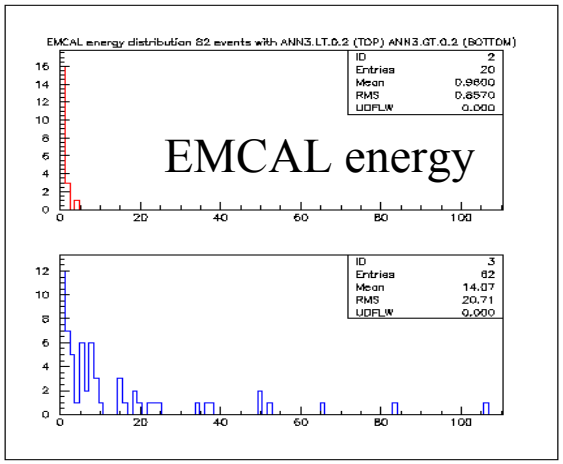


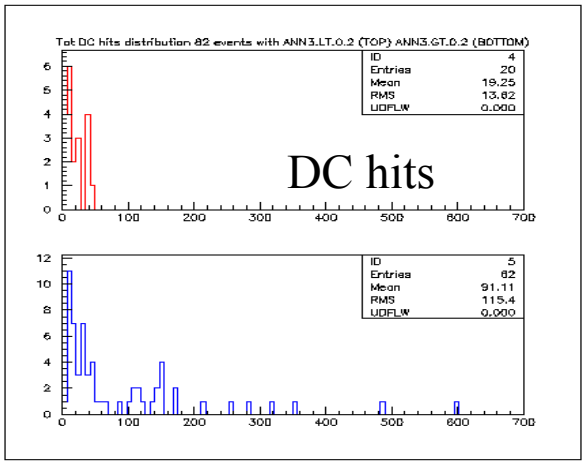
- The **distribution** of the output function of ANN 3 on the 82 selected neutrino events **should leave ~ 5 %** of the events (5 events) **below 0.2.**
- In this case **24 % is below 0.2** (20 events) and is far from what expected.
- **Either** some of these events are **not neutrino interactions** **or** they are **"new" to the ANN** (the **ANN** has been **trained** with **different neutrino events**)

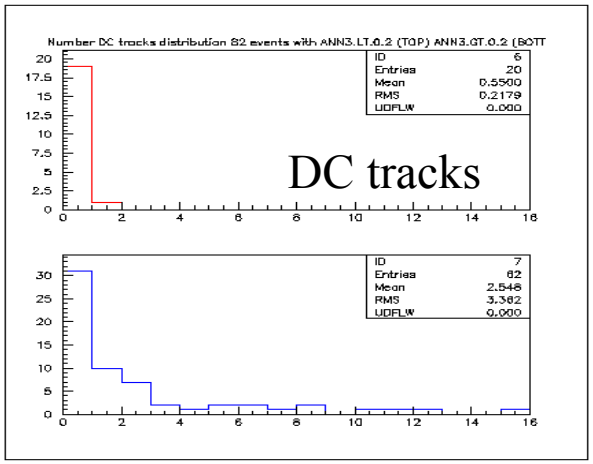
Characteristics of the 20 events below 0.2 cut from ANN3

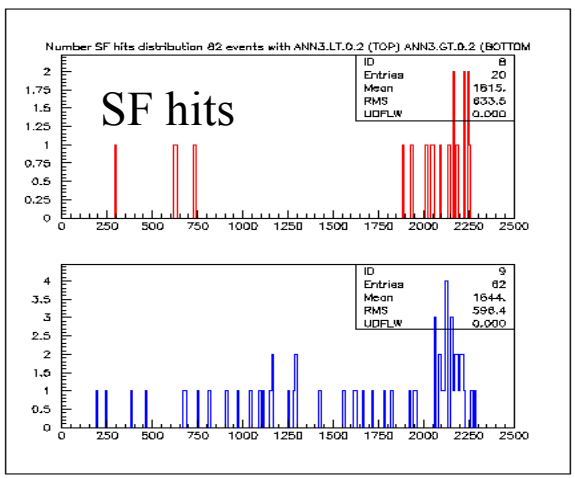
- Looking at these events with event display we observed that they all have the **following characteristics** (shown in the next distributions):
  - **Lots of activity in the SF 's** consistent with (most probably) shower initialization.
  - **No or very little energy** in the EMCAL
  - **No or very few DC hits** and **DC tracks**

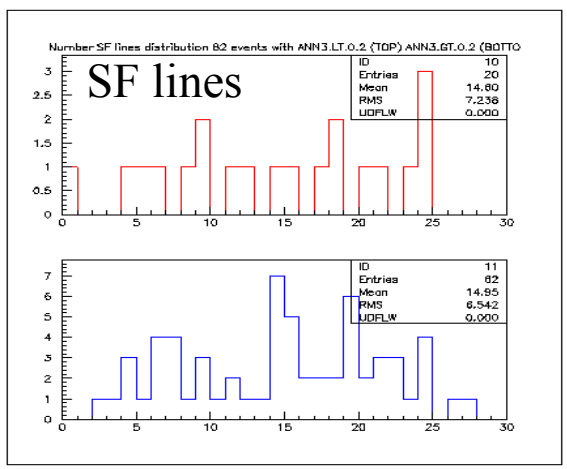
Characteristics of the 20 events below 0.2 cut from ANN3 (Distributions)



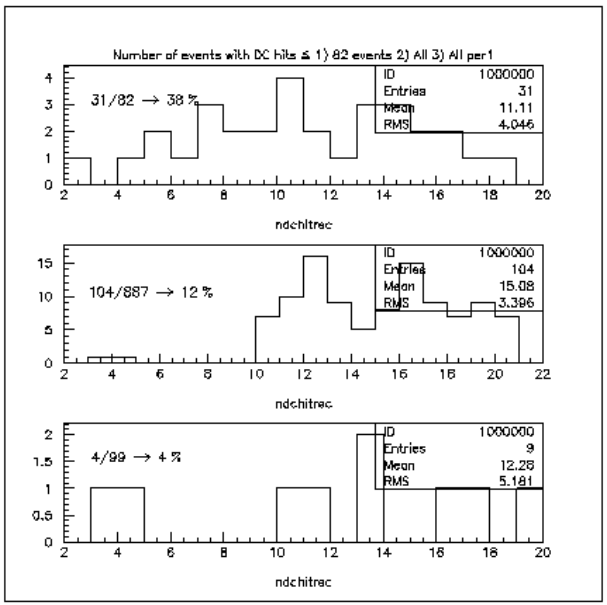


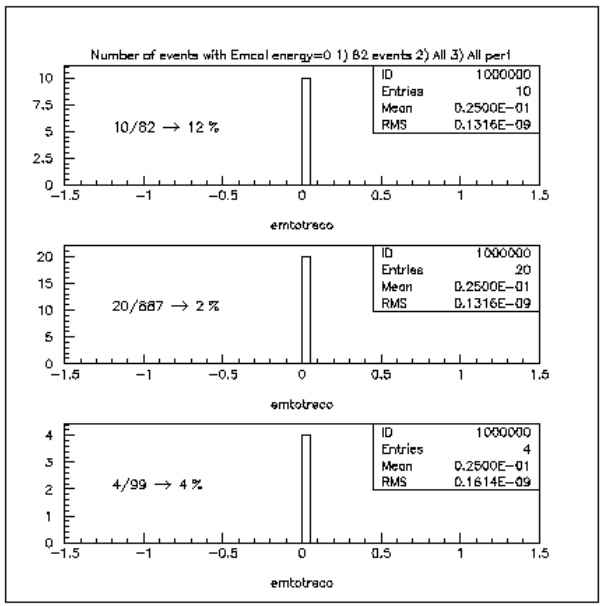






% of events with the previous characteristics on the 82 and 877 neutrino events





82 (per1) **877(all)** **99(of all in per1)**

% of events with EMCAL energy = 0

12 % **2%** **4%**

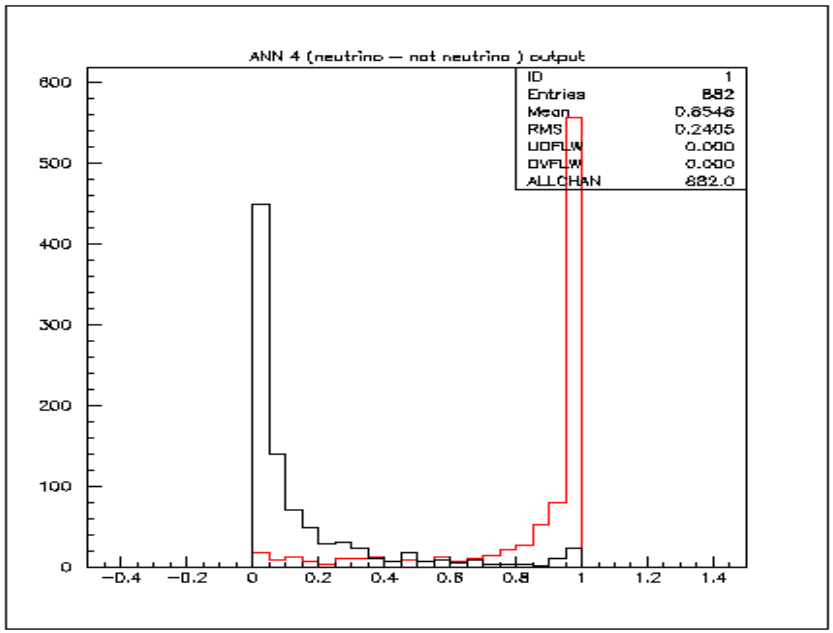
% of events with DC hits < 20

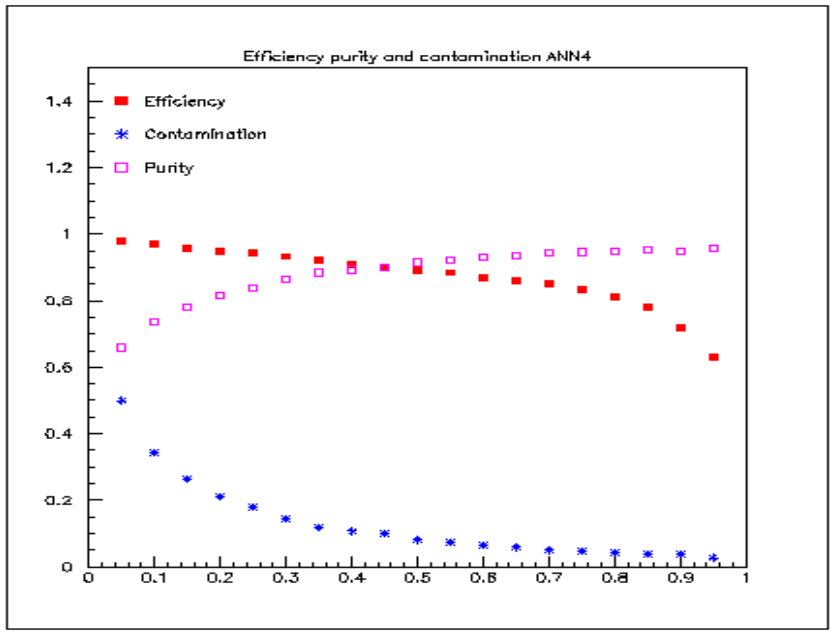
38 % **12%** **4%**

Construction of ANN 4 for neutrino event selection

- If during this scanning neutrino events are selected in a different way (selection criteria are changed) then ANN3 will not select such events since it is not trained to do so.
- Constructing ANN 4 we exclude variables related with EMCAL and DC info and add variables related with SF information as an attempt to account for these changes.
- Input variables of ANN 4:
  - Number of SF hits above 400 ph cut, Total Pulse height (ph cut)
  - % of hits and pulse height downstream the vertex, % of hits and pulse height upstream the vertex,% of "interaction" hits (ph cut), % of hits in each SF station, number of SF lines.
  - Number of VC hits,MID central hits, EMCAL energy along y=0 $|x| > 100$ cm.
  - T31,T21,T32

ANN 4 Output probability distributions

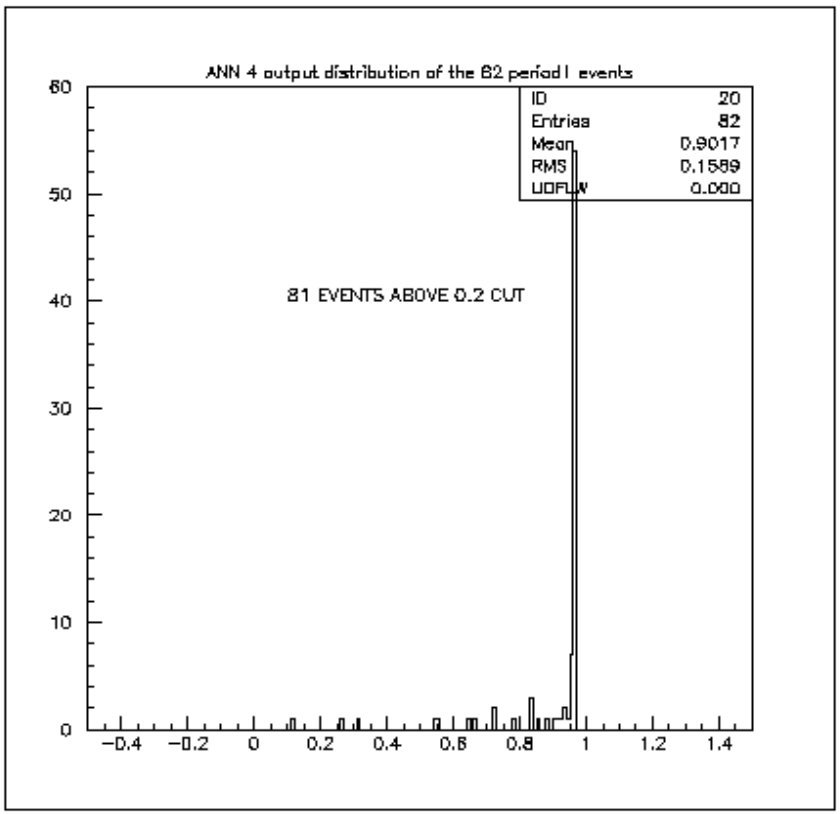




- With a **cut @ 0.2** in the network output function we select :

SIGNAL with **efficiency : 95 % purity 81% contamination 21%**

ANN4 output distribution on 82 per1 events



- **5 %** (4 events) of the 82 events is expected to be **below 0.2**
- **1.2 % (1 event)** is **below 0.2**.
- This **ANN4** seems **more efficient** in selecting neutrino events because by **construction** is **more tolerant**.

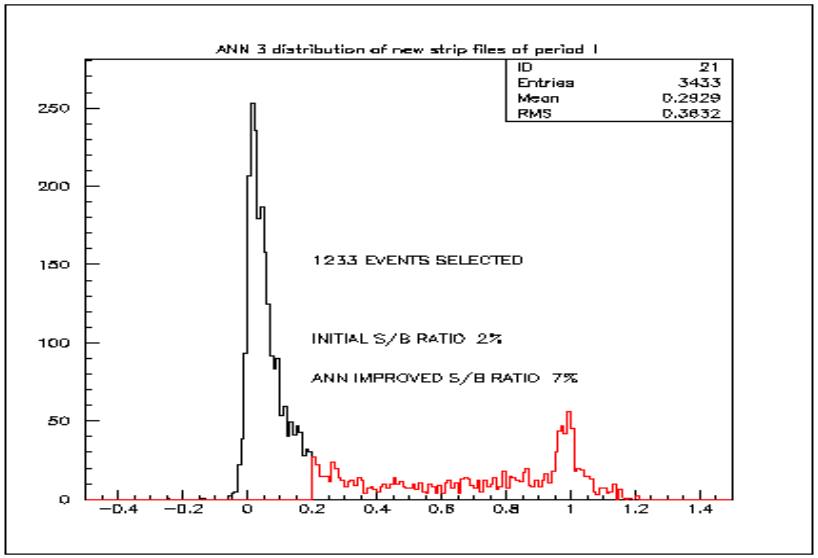
ANN and Signal/Background

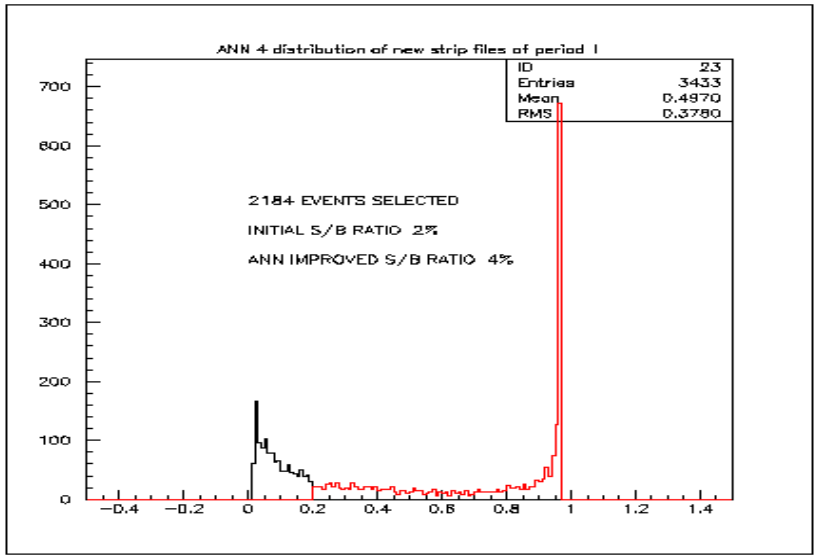
- We should also examine both **ANN's** taking into account the **Signal/Background** ratio **improvement.**
- In cases where the number of **"signal"** events is expected to be **very small** compared to that of the "background" the **Signal/Background improvement** after setting the cut must also be examined.
- **In the present case** :
  - **Signal events ~ 80, Background events ~ 3300** then **S/B ~ 2.4 %**
  - With a **cut @0.2 ANN3** should improve :

    S/B ~ 12 % (select 700 events from which 78 are neutrino)
  - With a **cut @0.2 ANN4** should improve :

    S/B ~ 10 % (select 780 events from which 78 are neutrino)

ANN3 & ANN4 - S/B on new strip files of per1





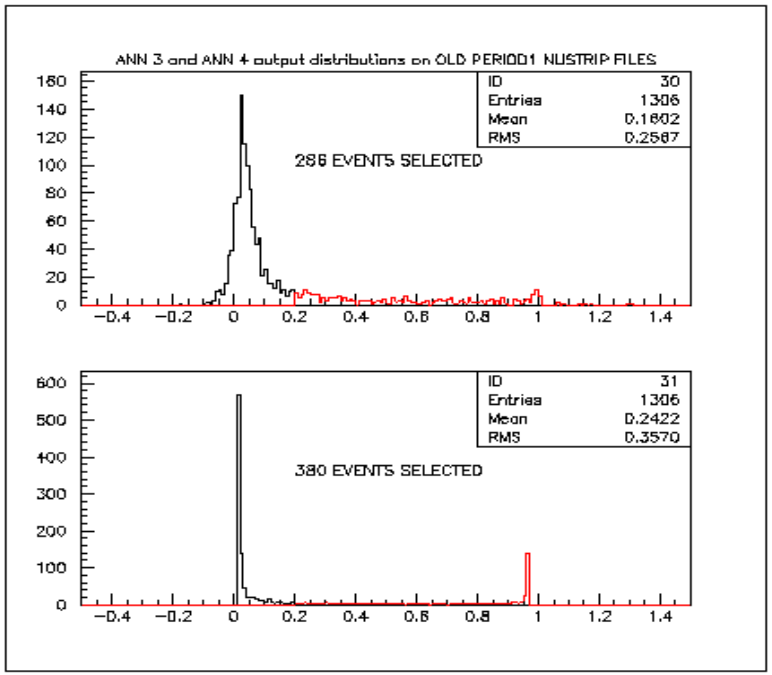
- The event sample consists of the **~ 3500** events in the new strip files of period 1 from which the **82 events** have been selected as **neutrino interactions.**

ANN3 - S/B ~ 5 % (selects 1233 events from which 62 are neutrino)

ANN4 - S/B ~ 4 % (selects 2184 events from which 81 are neutrino)

- Both are **lower** than what **expected** $\Rightarrow$ **Different criteria on selecting events to create the new strip files ?**

ANN3 & ANN4 - S/B on old strip files of per1



288/1310 events selected

380/1310 events selected

- Strip files 2907, 2911 2913,2929 **~ 1310 events & ~ 10 neutrino events**

ANN 3 $\frac{S}{B_{initial}} = 0.8\%, \quad \frac{S}{B_{expected}} = 4.5\%, \frac{S}{B_{observed}} = 3.5\%$

ANN 4 $\frac{S}{B_{initial}} = 0.8\%, \quad \frac{S}{B_{expected}} = 3.5\%, \frac{S}{B_{observed}} = 2.6\%$

Conclusions

- The possible change on the criteria with which "background" (strip files) and " signal" (neutrino) events are now selected affects the performance of both ANNs.
- As a result they are functioning below their capabilities.
- But both ANNs improve the signal to background ratio by a factor of 2 or even better and can be used anyhow.

  (**ANN3 selects 1233/3433** events in which and **62/82 neutrino** events are present & **ANN4 selects 2184/3433** in which **81/82 neutrino events** are present).
- If the criteria for "background" (strip files) selection stay unchanged for period 3 and 4, then the performance of the ANNs on these event samples will be even better.